DOCUMENT RESUME

ED 305 263 SE 050 477

AUTHOR Esquivel-Alfaro, Juan Manuel; Diaz-Solis, Sofia Maria

TITLE Science Achievement of Costa Rican Sixth Grade Students and Its Relationship with Selected Variables.

PUB DATE 89

NOTE 16p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (62nd, San Francisco, CA, March 30-April 1, 1989).

PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Academic Achievement; Achievement Tests; *Criterion Referenced Tests; *Elementary School Science; Foreign Countries; Grade 6; Intermediate Grades; *Science Tests

IDENTIFIERS *Costa Rica; *Science Achievement

ABSTRACT

This study had three purposes. First, to develop and validate a criterion-referenced test to measure science knowledge of students who finished the second cycle of the Basic General Education (6th grade). Second, to assess the performance of the entire Costa Rican population of sixth graders and, third, to analyze the results according to some selected variables. The tests were administered to the entire population of Costa Rican sixth graders. A total of 34,862 usable answer sheets were collected. Looking at the data by region, the same seven objectives were mastered by 50 percent or more of the students in 14 out of the 17 regions. Different results were obtained from comparing schools by size. Analyzing the means per objective, it can be concluded that the larger the school, the higher the means. (CW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED305263

THE UNIVERSITY OF COSTA RICA
SCHOOL OF EDUCATION
RESEARCH INSTITUTE FOR THE IMPROVEMENT OF
COSTA RICAN EDUCATION (I.I.M.E.C.)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SCIENCE ACHIEVEMENT OF COSTA RICAN SIXTH GRADE STUDENTS AND ITS RELATIONSHIP WITH SELECTED VARIABLES

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Juan M. Esquivel-Alfaro

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

JUAN MANUEL ESQUIVEL-ALFARO., Ph.D
BR. SOFIA MARIA DIAZ-SOLIS

Paper presented at the National Association for Research in Science Teaching, 1989 Annual Meeting at San Francisco, California

March 29 - April 1º

SE 050 477

BEST COPY AVAILABLE

SCIENCE ACHIEVEMENT OF COSTA RICAN SIXTH GRADE STUDENTS AND ITS RELATIONSHIP WITH SELECTED VARIABLES

Purpose

This study had three purposes. Firsts to develop and validate a criterion-referenced test to measure Science knowledge of students that finished the second cycle of the Basic General Education (6th grade). Second, to assess the performance of the entire Costa Rican population of sixth graders and third to analyze the results according to some selected variables.

Theoretical basis for study

The test developed in this study can be classified as a criterion-referenced test because it possesses the two characteristics which differenciate these tests from norm-referenced measurements instruments (Greco, 1974; Glaser and Nitko, 1971, Millman, 1980; Popham, 1978). First, it pretends to measure a very well defined and specific content domain and the results are interpreted in terms of the mastery of that content domain. Since criterion-referenced test procedures for measuring reliability are just emerging in the specialized literature, criteria for judging the procedure are still controversial. Procedures can be classified, according to Hambleton, et al (1978) and Subkoviac (1980), as squared-error loss methods, threshold-loss techniques or as decision-consistency indices. In an empirical study Dowing and Mehrens (1980) compared

several indices and concluded that KR-21 is useful when software for the more sofisticated indices is not available. It is also well established that the Kuder-Richardson Index is the lower bound to the C-R test's reliability (Glasnapp, Poggio and Miller, 1988).

Validity could be addressed with different types of information. Popham (1978) analyzed validity as descriptive, functional and domain-selection validity. Hambleton (1980) interpreted the validity issue as content and construct validity. Rose (1984) introduced the concept of instructional validity. On the other hand, Glasnapp, Poggio, and Miller (1988) established four types of validity information: content, curricular, instructional, and concurrent. In summary, it can be concluded that there is one aspect of paramount importance; the descriptive (content) validity in criterion-referenced testing.

Popham (1980), Berk (1978), Bornuth (1970), and Roid (1979) established several different forms to specify the content domain to be tested: behavioral objectives, item forms, amplified objectives, IOX specifications, item transformations, etc... Several item analysis statistics are presented in the literature: Berk (1980) reviewed and analyzed item- objective congruence, item discrimination, and item homogeneity indices.

Procedures

In order to specify and define the content domain of the test, a group of judges (twelve) were asked to select, from the national Science curriculum of the second cycle of the Basic General Educational graders (4th to 6 th), the 25 most basic and essential Science learnings for that cycle. These judges were experienced elementary Science supervisors and college Science educators.

The 18 more frequent skills chosen by the judges were translated by the researchers into behavioral objectives and these were converted into amplified objectives (Popham, 1978). (See Table 1). Experienced elementary teachers were paid to write multiple- choice items according to the rules established in the amplified objectives. A minimum of 12 items were constructed for each objective.

(Insert Table 1)

Separate list of the items and objectives were given to a group of ten judges Science educators and experienced elementary teachers and they were asked to pair each item with the objective it was supposed to measure. The acceptance level for each item was given by the correct matching of eigth or more judges. The procedures reviewed here provided evidence of the curricular and content validity of the test.

With all the accepted items several pilot parallel tests were assembled and field tested with groups of sixth graders in order to have a measure of item discrimination (Brennan, 1972) and difficulty, and to find whether the items were understood by the students or not. Items with discrimination indices over 0.30 were accepted for the item bank.

With the best items, two 48-item tests were constructed, each measuring sixteen objectives. Each objective, therefore, was measured by three items. Fourteen ot the sixteen objectives were common to both tests.

The test cut-score standards were set using acceptable procedures that require judgement from experienced teachers and supervisors (Hambleton, 1980). The cut-score was set at two for all objectives. Scores from 0 to 1.99 mean no-mastery of the objective, scores of 2 and over mean mastery of a particular objective. Mastery scores and percentage of students who pass the objective are presented in the analysis of the data.

The tests were administered almost to the entire population of sixth graders in Costa Rica. The test were given to 40,740 students. Students in one-room schools were not tested because of difficulties that were raised from differences in the curriculum and hard-to-reach school buildings.

Results

The internal consistency of the tests calculated using Kuder-Richardson-21 procedure were: 0.83 and 0.85. The item congruency indices range from 0.30 to 0.94, but 79% of the items have discrimination indices over 0.50 (See Table 2).

(Insert Table 2)

A total of 34,862 answer sheets were usable for the analysis. The data were analyzed in order to compare mastery of the 18 objectives by: a) urban and rural schools, b) private and official schools, c) schools of different size and d) schools of the country's different regions.

At the national level the mean per objective varies from a minimum of 0.97 (objective 16) to a maximum of 2.85 (objective 18). Ten of the eighteen objectives have means over two (the cut-off score). The percentage of students mastering any particular objective runs from a five percent to an 88% Seven out of the 18 objectives are mastered by more than 50 percent of the students. This data can be observed in Table 3.

(Insert Table 3)

Looking at the data by region, the same seven objectives are mastered by 50 percent or more of the students in fourteen out of the 17 regions.

In Costa Rica, school size is a classification made according to the number of teachers working in a school building. Besides the one-room school, there are five schools sizes. Number one is the school with more than one teacher but less than six teachers; number two have from six to 10 teachers; number three from 11 to 20 teachers, number four from 21 to 30 teachers, and number five has more than 30 teachers. The results obtained by schools size showed that the same ten objectives are mastered by the schools of different sizes, with the exception of objectives four and 10, as can be observed in table 4.

(Insert Table 4)

The achievement of the private schools is better than that of the public schools. In Table 5 it can be observed that in the private schools 14 objectives were mastered by the students, while 10 objectives had means over two in the public schools. Ten objectives are mastered by 50% or more of the students in the private sector compared with seven in the public sector.

(Insert Table 5)

The urban schools did nor differ from the rural schools in the number of objectives mastered by 50% or more of the students. Ten objectives had means over two in the urban schools and nine objectives presented that same condition in the rural schools. (See Table 6).

(Insert Table 6)

Analyzing the value of the means per objective in general, it can be concluded that the larger the school, the higher the mean. The same trend holds when the urban schools are compared with the rural ones, and when private schools are compared with rural schools.

Conclusions

The test developed for this study showed a high reliability index. The procedures employed to define the content domain, to measure the item-objective congruency and to select the items, were evidence for the curricular and descriptive validity of the test.

In general, Science achievement in sixth grade is good since most of the objectives are mastered by the national population. It is much higher than the achievement showed in Mathematics, Spanish, and Social Studies (Delgado et al, 1988) by the same population.

As in several other studies (Esquivel and Quesada, 1985; Esquivel, Peralta and Delgado, 1983), there are differences in achievement between the private and public schools; between the urban and rural schools; and between schools of different sizes. Such differences are confirmed by this particular study.

Bibliography

Berk, R. A. (1980) Item analysis. In R. A. Berk (Ed) Criterion- referenced Measurement: The state of the art. Baltimore, Maryland: The Johns Hopkins University Press.

Berk, R.A. (1979) A Critical Review of content domain specification/item generation strategies for criterion-referenced tests. Paper presented at the anual meeting of the American Education Research association, San Francisco, (ERIC Document Reproduction Service Nº Ed 170382).

Bormuth, J. R. (1970) On the theory of achievement test itemes. Chicago: University of Chicago Press.

Brennan, R.L. (1972) A generalized upper-lower item discrimination index. Educational and psychological Measurement, 32, 289-303.

Dowing, S.M. y Mehrens, W. A. (1979). Six single - administration reliability coefficients for criterion - referenced tests: A comparative study. Paper presented at the annual meeting of AERA. Ontario, 1979.

Esquivel, J. M; Peralta, T. y Delgado, V. (1983). Desarrollo y validacion de pruebas de conocimientos minimos en Matemática y su aplicación a una muestra nacional de escuelas y colegios. Revista Universidad de Costa Rica: Educación 7 (1 y 2), P. 125-134.

Esquivel, J. M. y Quesada, L. (1985). The development and validation and administration of a Criterion-referenced Science Battery. Presented at the annual convention of NARST, French Licks Spring, Indiana, April 15-18.

Glaser, R. and Nitko, A. J. (1970). Measurement in learning instruction. In R. L Thorndike (Ed), Educational Measurement. (2ª Ed). Washington: American Council on Education.

Glassnap, D., Poggio, J., and Miller, M;. (1988) Kansas. Minimum Compentency testing program: Research and evaluation results 1980-1987. Lawrence, Ks. Center of Educational Testing and Evaluation. The University of Kansas.

Greco. T.H. (1974) Is there really a difference between criterion-referenced and norm-referenced measurement. Educational Technology. 22-25.

Hambleton, R. K. (1980). Test score validity and standard - setting methods. In R. A. Berk (Ed) Criterion - referenced Measurement: The State of the art. Baltimore, Maryland: The John Hopkins University Press.

Hambleton, R. K., et al (1978). Criterion-referenced testing and measurement: a review of technical issues and developments. Review of Educational Research, 48, 1-48.

Millman, J. 1980). Computer-based item generation. In R. A. Berk (Ed). Criterion-referenced Measurement: The state of the art. Baltimore: The Johns Hopkins University Press.

Popham, W. J. (1978). Criterion-referenced Measurement. Englewood Cliff, New Yersey: Prentice-Hall, Inc.

Popham, W. J. (1980) Domain specification strategies. In R.A. Berk (Ed) Criterion-referenced Measurement: The state of the art. Baltimore, Maryland. The Johns Hopkins University Press.

Roid, G. H. (1979) The technology of test - item writing. In H.F. O'Neil Jr. (Ed) Procedures for instructional system development. New York. Academic Press.

Rose, J, et al. (1984) Instructional validity: Merging curricular, instructional and test development issues. Paper presented at the annual meeting of the AERA, New Orleans.

Sukoviac, M (1980) Decisions-making approaches. . In R.A. Berk (Ed) Criterion-referenced Measurement: The state of the art. Baltimore, Maryland. The Johns Hopkins University Press.

Table 1

Objectives measured by the test

1.- Given a set of drawings, to identify the main parts of plants

2.- To identifiy the main functions of flowers, roots, stem, leaves and fruits.

3.- To differentiate between vertebrate and invertebrate animals.

4.- To identifiy the five classes of vertebrates according with their main characteristics.

5.- Given a paragraph , to identify a case of air or water pollution.

6.- Given a paragraph, to recognize appliances which produce caloric energy.

7.- To recognize the consequences of the earth's movements: rotation and translation.

8.- Given a set of drawings, to identifiy cells, tissues, organs and systems.

9.- To recognize the basic food groups: proteins, carbohydrates and fats.

10.- Given a paragraph, to identify harmful situations for one's health on food consumption and manipulation.

11.- Given a paragraph, to identify harmful organisms which could contaminate food.

12.- Given a set of drawings, to recognize productors, consumers and decomposers organisms

13.- To recognize parasite organisms.

14.- To relate parts of the human reproductive system with the human sex (masculine and femenine).

15.- To relate components of the masculine and femenine reproductive systems with their function.

16.- To identify sedimentary and igneous rocks according to their origin and characteristics.

17.- Given a set of drawings, to identify electricity conductors.

18.- To identify electric appliances.

TABLE 2

Discrimination index (Breman) frecuency distribution

D.I. Range	f	%
0.30 - 0.39	8	8.3
0.40 - 0.49	12	12.5
0.50 - 0.59	22	22.9
0.60 - 0.69	26	27.1
0.70 - 0.79	13	13.5
0.80 - 0.89	10	10.4
0.90 - 0.99	5	5.2
TOTAL	96	100.0

TABLE 3

Objective means and percentage of students mastering each objective

OBJECTIVE	X	%	OBJECTIVE	X	%
1	2.27*	81.0	10	2.04*	40.4
2	1.18	28.8	11	2.73*	78.5
3	2.37*	59.8	12	1.92	29.2
4	2.04*	35.1	13	2.32*	53.8
5	1.32	14.8	14	2.54*	68.4
6	2.50*	63.9	15	1.88	32.8
7	1.44	20.8	16	0.97	5.0
8	2.17*	43.8	17	1.93	36.7
9	1.82	18.8	18	2.85*	88.0

* Values over the cut-off score

TABLE 4

Objective means by school size

OBJECTIVE	SCHOOL SIZE				
	1	2	3	4	5
1	2.73	2.77	2.78	2.77	2.78
2	1.81	1.79	1.79	1.74	1.78
3	2.32	2.31	2.35	2.36	2.41
4	2.06	2.03	2.00	1.98	1.24 ←
5	1.44	1.40	1.26	1.24	1.28
6	2.52	2.47	2.47	2.46	2.51
7	1.40	1.34	1.37	1.33	1.47
8	2.03	2.10	2.16	2.16	2.25
9	1.87	1.84	1.79	1.78	1.81
10	1.93	1.98	2.02	2.01	2.09 ←
11	2.72	2.70	2.72	2.71	2.74
12	1.87	1.88	1.87	1.87	1.95
13	2.40	2.29	2.33	2.27	2.30
14	2.36	2.47	2.54	2.59	2.62
15	1.74	1.80	1.86	1.88	1.94
16	0.97	0.96	0.92	0.93	0.96
17	1.96	1.88	1.85	1.86	1.94
18	2.82	2.85	2.85	2.86	2.86
N	8,211	2,593	5,008	5,063	11,785

TABLE 5

OBJECTIVE MEANS AND PERCENTAGE OF STUDENTS MASTERING EACH OBJECTIVE IN PRIVATE AND PUBLIC SCHOOLS

OBJECTIVE	PRIVATE SCHOOLS		PUBLIC SCHOOLS		OBJECTIVE	PRIVATE SCHOOLS		PUBLIC SCHOOLS	
	X	%	X	%		X	%	X	%
1	2.82	84.7	2.77	80.8	10	2.77	81.8	2.72	78.2
2	2.12	44.9	1.79	27.9	11	2.33	56.1	2.02	39.5
3	2.53	68.7	2.36	59.3	12	2.23	45.6	1.90	28.2
4	2.33	49.5	2.03	34.2	13	2.34	53.0	2.32	53.7
5	1.34	14.9	1.32	14.6	14	2.74	79.6	2.53	67.8
6	2.66	73.4	2.49	63.4	15	2.16	45.8	1.86	32.0
7	1.97	44.0	1.41	19.5	16	1.17	7.8	0.95	4.8
8	2.45	57.8	2.16	43.0	17	2.30	54.0	1.91	35.7
9	1.82	20.1	1.87	18.7	18	2.89	91.0	2.85	87.8

TABLE 6

OBJECTIVE MEANS AND PERCENTAGE OF STUDENTS MASTERING EACH OBJECTIVE IN URBAN AND RURAL SCHOOLS

OBJECTIVE	RURAL		URBAN		OBJECTIVE	RURAL		URBAN	
	X	%	X	%		X	%	X	%
1	2.75	79.5	2.74	82.3	10	1.96	37.3	2.10	43.0
2	1.79	27.2	1.82	30.2	11	2.71	77.5	2.74	79.2
3	2.31	57.2	2.42	62.0	12	1.86	27.2	1.86	30.7
4	2.03	34.1	2.05	35.8	13	2.34	55.0	2.31	52.7
5	1.36	15.9	1.29	13.8	14	2.43	62.7	2.63	73.1
6	2.49	63.6	2.51	64.2	15	1.79	28.6	1.95	36.2
7	1.38	23.2	1.48	17.8	16	0.93	4.8	0.99	5.2
8	2.08	39.3	2.25	47.5	17	1.91	35.5	1.95	37.7
9	1.84	20.6	1.81	17.3	18	2.84	86.7	2.87	89.1